

Our Ref.: 80398.P322

UNITED STATES PATENT APPLICATION

FOR

SYSTEM AND METHOD FOR HUMAN FACE DETECTION  
IN COLOR GRAPHICS IMAGES

Inventor: Lizhi Wang

Prepared By:

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN  
12400 Wilshire Blvd., 7th Floor  
Los Angeles, California 90025-1026  
(310) 207-3800

EXPRESS MAIL CERTIFICATE OF MAILING

I hereby certify that this correspondence is being  
deposited with the United States Postal Service as  
Express Mail No. EL 234 217 792 US  
addressed to: Commissioner of Patents and  
Trademarks, Washington, D.C. 20231, on

June 2, 2000

Date

Cheri Clinkenbeard

Typed or printed name of person mailing paper or fee

Signature of person mailing paper or fee

002090-08658560

## BACKGROUND OF THE INVENTION

### (1) Field of the Invention

The present invention relates to the field of computer graphics. In particular, the present invention relates to a system and a method for detecting human faces in an existing color graphics image.

### (2) Description of Related Art

An automatic face recognition system should have the ability to identify one or more persons in a scene, starting from still or video images of the scene.

A complete solution of the problem involves segmentation of faces from cluttered scenes, extraction of features from the face region, identification and matching. The first step in this process is face detection. The goal of face detection is to determine whether there is one or more human faces in the image, and, if present, return its location and spatial extent for further processing.

Most of the existing face detection systems use window-based or pixel-based operation to detect faces. In a window-based system, a small window is moved over all portions of an image to determine whether a face exists in each window based on distance metrics. Common problems with window-based approaches are that they cannot detect faces of different orientations or view angles, and that they are computationally expensive.

In representative pixel-based analyses, color segmentation is used as the first step. This assumes that human skin colors fall in a small, known region in color space. Some advantages of using color as feature are speed and invariance to orientation. However, it is difficult to model human skin color for several reasons. One reason is that different cameras produce significantly



## SUMMARY

A system and method for determining a series of candidate patches for human faces in a color graphic image is disclosed. The method may start by  
5 determining a first area wherein a color gradient has a low value, and then determining a second area wherein an intensity value has a high value. The method may then perform a logical AND on the first area and the second area to create a third area. The method may then select portions of this third area with suitable hue saturation values to form a series of candidate patches where  
10 the human faces reside.

002090" 0865560

## BRIEF DESCRIPTION OF THE DRAWINGS

Fig 1

The features, aspects, and advantages of the present invention will become more fully apparent from the following detailed description, appended claims, and accompanying drawings in which:

Figure 1 is a block diagram of a computer graphics system, according to one embodiment of the present invention;

Figure 2 is an unprocessed graphical image, according to one embodiment of the present invention;

Figure 3 is a color gradient map of the image of Figure 2, according to one embodiment of the present invention;

Figure 4 is an intensity map of the image of Figure 2, according to one embodiment of the present invention;

Figure 5 is a combined map of the color gradient map of Figure 3 and the intensity map of Figure 4, according to one embodiment of the present invention;

Figure 6 is a hue spectrum chart, according to one embodiment of the present invention;

Figure 7 is a set of candidate patches and determined ellipses for the image of Figure 2, according to one embodiment of the present invention;

Figure 8 shows the face candidate map superimposed on the image of Figure 2, according to one embodiment of the present invention;

Figure 9 shows the final result of the process, according to one embodiment of the present invention; and

Figure 10 is a flowchart of process steps, according to one embodiment of the present invention.

## DETAILED DESCRIPTION

In the following description, numerous specific details are set forth to provide a thorough understanding of the present invention. However, one  
5 having an ordinary skill in the art may be able to practice the invention without these specific details. In some instances, well-known circuits, structures, and techniques have not been shown in detail in order not to unnecessarily obscure the present invention.

Psychological studies of human face recognition by humans suggest that  
10 virtually every kind of available information is used simultaneously. For example, the configuration of the facial features can help humans find the location of the face, since it is known that the features can not appear in arbitrary arrangements. In contrast, a computer graphics system cannot use all kinds of information simultaneously. A limited number of kinds of  
15 information must be processed at any given time. Important constraints to the processing of selected kinds of information include simplicity of estimation of parameters, low dependency upon ambient light intensity, low dependency upon small changes of facial expression, and maximizing information content in those kinds of information selected.

20 In one embodiment of the present invention, color, shape, intensity, and face configuration information are combined together to locate faces in color images. This method of face detection advantageously is computationally efficient.

Referring now to Figure 1, a block diagram of a computer graphics  
25 system 100 is shown, according to one embodiment of the present invention. Computer graphics system 100 may include read-only memory (ROM) 102, random-access memory (RAM) 104, one or more central processing units (CPU)

106, one or more graphics controllers 108 with attached displays 110,  
connections to a local area network (LAN) 120 and wide area network (WAN)  
122 via LAN controller 112 and WAN controller 114, respectively, mass storage  
devices 116, and removable media 118. The functional parts of computer  
5 graphics system 100 may be connected via one or more system busses 124.  
Software executing on computer graphics system 100 may be loaded via the  
LAN 120, the WAN 122, or the removable media 118.

Referring now to Figure 2, an unprocessed graphical image 200 is shown,  
according to one embodiment of the present invention. Exemplary graphical  
10 image 200 includes five faces, 202, 204, 206, 208, and 210. Detecting these  
faces and exhibiting their size, shape, and location will be discussed in  
conjunction with Figure 2 and related figures, Figures 3, 4, 5, 7, 8, and 9  
below. In one embodiment, the following four assumptions are made: the  
facial area is relatively smooth and bright; the only lighting is ambient  
15 illumination; the outlines of human faces are roughly elliptical; and there are  
significant color and intensity changes inside facial areas because of the  
configuration of facial features.

The image of Figure 2 is represented by a large number of picture  
elements (pixels), each with three words of data representing color and  
20 intensity. The words in one embodiment may include 8 bits. In alternate  
embodiments other word lengths may be used. In one embodiment, the three  
words may correspond to red intensity, green intensity, and blue intensity  
(RGB format). In alternate embodiments, the three words may be separated in  
other color spaces, such as luminance, chrominance minus blue, chrominance  
25 minus red (YCC format).

Referring now to Figure 3, a color gradient map 300 of the image of  
Figure 2 is shown, according to one embodiment of the present invention. In

one embodiment, the color gradient is the rate of change of color when going from one area to another. The color gradient may be used to extract homogeneous regions since it may be assumed that the facial area is relatively smooth. The color gradient may also extract edge information.

5 In order to calculate the color gradient, various methods of calculation may be used. In one embodiment, a Sobel filter, well-known in the art, is used to calculate a map of the magnitude of the color gradient. Commercially available Sobel filter software is available for many common computer graphics workstations.

10 The map of the magnitude of the color gradient, determined immediately above, is then converted by applying a true or false threshold value into a map wherein each pixel is represented by a single bit of data, true or false. The threshold is determined by the process of normalization. In normalization, the average value of the magnitude of the color gradient of all the pixels is  
15 determined. A fixed percentage of the average value is selected as the threshold, and then the magnitude of the color gradient for each pixel is compared with the threshold. If the magnitude is less than the threshold, the pixel is considered a "true". If the magnitude is more than the threshold, the pixel is considered a "false".

20 In one embodiment, the map wherein each pixel is represented by a single bit of data, determined as in the paragraph immediately above, is converted into a simplified mosaic representation. Use of a simplified mosaic representation filters out any noise yet retains the facial information. In one embodiment, each group of 4 by 4 pixels is replaced by a single mosaic pixel  
25 with a value assigned by determining the predominant value, true or false, assigned to the 16 pixels within the mosaic pixel. In other embodiments, other

sizes of mosaic blocks may be used. This simplified mosaic representation of the above-described map may be termed a color gradient map.

Figure 3 is a representative color gradient map for the image of Figure 2, derived in accordance with the steps outlined above. The white area 310 is a region wherein the mosaic pixel values are all true. The black area 320 is a region wherein the mosaic pixel values are all false. In accordance with one embodiment, the black area 320 are areas rejected as candidate face locations, whereas the white areas 310 are areas considered for further processing as candidate face locations.

Referring now to Figure 4, an intensity map 400 of the image of Figure 2 is shown, according to one embodiment of the present invention. It is assumed that any faces in the image are well illuminated and well focused. Therefore, the facial areas may have higher intensities than the surrounding areas.

Again, a threshold value may be determined by the process of normalization. An intensity value for each pixel is first determined. An average value of the intensity values for all the pixels is then determined. A percentage of this average value is then selected as a threshold value.

The intensity value of each pixel is then compared with the threshold value. If the intensity of a pixel is greater than the threshold value, the pixel is assigned a value "true". If the intensity of a pixel is lower than the threshold value, the pixel is assigned a value "false".

As in the case of the color gradient map derivation, in one embodiment the map wherein each pixel is represented by a single bit of data, determined as in the paragraph immediately above, is converted into a simplified mosaic representation. In one embodiment, each group of 4 by 4 pixels is replaced by a single mosaic pixel with a value assigned by determining the predominant value, true or false, assigned to the 16 pixels within the mosaic pixel. In other

embodiments, other sizes of mosaic blocks may be used. This simplified mosaic representation may be termed an intensity map.

Figure 4 is a representative intensity map for the image of Figure 2, derived in accordance with the steps outlined above. The white area 410 is a region wherein the mosaic pixel values are all true. The black area 420 is a region wherein the mosaic pixel values are all false. In accordance with one embodiment, the black area 420 are areas rejected as candidate face locations, whereas the white areas 410 are areas considered for further processing as candidate face locations.

Referring now to Figure 5, a combined map 500 of the color gradient map of Figure 3 and the intensity map of Figure 4 is shown, according to one embodiment of the present invention. Each mosaic pixel of a combined map is first determined to be either true or false by performing a binary AND operation on the respective mosaic pixels in the color gradient map and the intensity map. Those resulting mosaic pixels whose value is true are areas that combine smoothness and brightness.

After performing the logical AND operation, a process known as morphological erosion is performed on the resulting map. Morphological erosion removes small true areas surrounding predominantly true areas in the following manner. For each mosaic pixel that has a value of true, examine all the neighboring mosaic pixels. If any of the neighboring mosaic pixels have a value of false, then replace the value of the central mosaic pixel with the value false. When this operation is performed a single time for all originally true pixels, the result is a map wherein each true area has had its peripheral area eroded. Sufficiently small true areas are removed entirely.

Figure 5 shows the resulting combination map formed by the AND operation performed on Figure 3 and Figure 4, followed by one pass of the

morphological erosion process. The white area 510 shows where the value is true, signifying candidate face locations. The black area 520 shows areas rejected from further consideration as face locations.

Referring now to Figure 6, a hue spectrum chart is shown, according to one embodiment of the present invention. Human skin color has been used as an important cue to locate human face in color images, based on the research results that human skin color tends to cluster in a pair of compact regions in certain transformed color space. One such transformed color space is hue saturation intensity (HSI) space. The chart of Figure 6 shows a range in HSI space. In HSI space, region 610 is reddish-green, regions 620, 630 are greenish, region 640 is bluish, region 650 is purplish, and region 660 is bluish-red.

Referring now to Figure 7, a set of candidate patches and determined ellipses for the image of Figure 2 is shown, according to one embodiment of the present invention. The image of Figure 2 is first converted to HSI space, using commercially available software. The candidate patches of Figure 7 are then created by determining the parts of white area 510 of Figure 5 which have HSI space values of region 610 and region 660. The resulting candidate patches 710, 720, 730, 740, and 750 are shown as irregular patches in Figure 7.

Another step of morphological erosion may be performed on each candidate patch. As in the discussion in connection with Figure 5, each mosaic pixel in a candidate patch is examined to see if any neighboring mosaic pixels are not within the candidate patch. If so, then the examined mosaic pixel is removed from the candidate patch.

Human facial outlines are consistently roughly elliptical. Therefore, for each candidate patch 710, 720, 730, 740, and 750 a corresponding ellipse 712, 722, 732, 742, and 752, respectively, is determined. In one embodiment, a

Hoteling transform is used to create the corresponding ellipses. Commercially available Hoteling transform software may be used. In this step, the principle axis (major axis and minor axis) for each candidate patch are determined. The ellipse is completely specified by the principle axis.

5        Next the degree of fit of the ellipse is measured. In one embodiment, the degree of fit may be measured by counting the number of mosaic pixels of the candidate patch that are within the boundaries of the corresponding ellipse. The ratio of the number of mosaic pixels of the candidate patch within the corresponding ellipse to the total number of mosaic pixels in the candidate  
10       patch may be used as a measure of degree of fit. In one embodiment, if the above ratio is above 80%, the degree of fit is said to be good; if the above ratio is below 80%, the degree of fit is said to be bad. In alternate embodiments, other measurements for good and bad fit may be used.

Referring now to Figure 8, the face candidate map is superimposed on  
15       the image of Figure 2, according to one embodiment of the present invention. Putting the face candidate map of Figure 7 on the top of the original color image of Figure 2 yields Figure 8 which shows the detection initial result.

Candidate patches with bad degrees of fit may be further subdivided. For example, candidate patch 720 includes faces 202 and 204. In order to  
20       further subdivide candidate patch 720, the above process of deriving the color gradient map and intensity map are repeated within the candidate patch. A new normalization for thresholding may be performed. New candidate patches may be identified, and each of these may have a new ellipse calculated. If the new ellipses are again bad fits, then the process of subdividing may be  
25       repeated as necessary.

Whether or not the candidate patches are further subdivided, they are finally examined for a lack of detail. Human faces have great variety within the

overall shape because of the features, which may include the eyes, nose, and mouth. Each candidate patch may therefore be examined by calculating the standard deviation of intensity within each candidate patch. If the intensity within the candidate patch exceeds a standard deviation from the mean value at certain places, the candidate patch is considered sufficiently "face like". If, however, the intensity fails to exceed a standard deviation from the mean value, then the candidate patch is removed from further consideration. As an example, the star shape at candidate patch 710 fails this particular test, so it is removed from further consideration.

Referring now to Figure 9, the final result of the process is shown, according to one embodiment of the present invention. Comparing previous and final results, we can see that the largest candidate patch 720 has been split into two candidate patches 924, 920, and the golden star at candidate patch 710 has been filtered out. The process has successfully and efficiently identified faces 202, 204, 206, and 208 of Figure 2. Face 210 was missed due to lack of intensity and color gradient changes. Hand 222 was mistakenly identified as a face.

Referring now to Figure 10, a flowchart of process steps is shown, according to one embodiment of the present invention. The process begins at start 1000 with the selection of a color graphics image. In two parallel steps, smoothness detection 1010 and brightness detection 1020, the color gradient map (discussed in connection with Figure 3 above) and the intensity map (discussed in connection with Figure 4 above) are derived. Then in the binary AND 1030 step, the color gradient map and the intensity map are combined.

In step 1032 a step of morphological erosion is performed. Then in step 1034 a hue detection step is performed, yielding initial candidate patches. After a second step of morphological erosion 1036 is performed, the ellipses for

each candidate patch are determined in an ellipse fitting 1038 step. Then in step 1040, it is determined whether or not the ellipses are a good fit. If not, then the candidate patch is subjected to a further re-evaluation step 1042. After a further morphological erosion step 1044, a subsequent step of ellipse fitting 1046 is performed. These subsequent ellipses are again determined to be a good fit or a bad fit in step 1040.

If, however, the ellipses are determined to be a good fit to the candidate patches in step 1040, then a subsequent determination is made whether the candidate patch is too smooth (e.g. lacks detail) in step 1050. Step 1050 may include the variation by more than a standard deviation approach used in connection with Figure 8 above. If the candidate patch is determined to be too smooth to be a face, then in step 1054 the candidate patch is considered a bad patch and discarded. If, however, the candidate patch is determined to contain suitable variation, then in step 1052 the candidate patch is added to the list of good candidate patches. The Figure 10 process then ends 1060 with a list of good candidate patches.

In the foregoing specification, the invention has been described with reference to specific embodiments thereof. It will however be evident that various modifications and changes can be made thereto without departing from the broader spirit and scope of the invention as set forth in the appended claims. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense. Therefore, the scope of the invention should be limited only by the appended claims.